Sampling Methods for Web and E-mail Surveys

Ronald D. Fricker, Jr.
Naval Postgraduate School
9 October 2006

Abstract. This chapter is a comprehensive overview of sampling methods for web and e-mail ('Internet-based') surveys. It reviews the various types of sampling methods – both probability and non-probability – and examines their applicability to Internet-based surveys. Issues related to Internet-based survey sampling are discussed, including difficulties assembling sampling frames for probability sampling, coverage issues, and nonresponse and selection bias. The implications of the various survey mode choices on statistical inference and analyses are summarized.

DRAFT

Introduction

In the context of conducting surveys or collecting data, sampling is the selection of some subset of a larger population. As this chapter discusses, there are various ways by which the selection may be accomplished and the resulting sample may include a small or large fraction of the population. In fact, it can include every member of the population. A survey based on such a sample is called a *census*.

In this chapter we focus on sampling methods for web-based and e-mail surveys, which taken together we call 'Internet-based surveys'. In our discussion we will frequently compare sampling methods for Internet-based surveys to various types of non-Internet-based surveys, such as postal mail and the telephone surveys, which in the aggregate we refer to as 'traditional surveys'.

In this chapter we begin with a general overview of sampling, including the various sampling methods and the basic terminology. Since there are many fine textbooks on the mechanics and mathematics of sampling, we restrict our discussion to the main points and issues that will be necessary to ground our discussion on sampling for Internet-based surveys. We then discuss the various types of Internet-based surveys and the sampling methods applicable to each

followed by an examination of the issues and challenges associated with Internet-based surveys.

Throughout the chapter the discussion predominantly focuses on surveys, but surveys are just a means for collecting data. The sampling methods we discuss, as well as the issues and challenges associated with conducting Internet-based surveys, broadly apply to other types of data collection conducted via the Internet and hence are relevant to other types of research as well. In addition, we briefly discuss issues related to nonresponse, solicitation of participation, and incentives for Internet-based surveys.

Why Sample?

Surveys are often conducted to gather information about a population. Sometimes the survey is conducted as a census, where the goal is to survey every unit in the population. However, it is frequently impractical or impossible to survey an entire population, perhaps due either to cost constraints or some other practical constraints, such as it may not be possible to identify all the members of the population.

An alternative to conducting a census survey is to select a sample from the population and only survey those sampled units. A survey administered to an appropriately selected sample can have a number of advantages over a census, including:

- Lower cost
- Less effort to administer
- Better response rates
- Greater accuracy

The advantages of lower cost and less effort are probably obvious. Keeping all else constant, reducing the number of surveys should cost less and take less effort to field and analyze. However, that a survey based on a sample rather than a census can give better response rates and greater accuracy is less obvious.

DRAFT

Better response rates can result when, for a fixed level of effort (or funding), a sample allows the surveying organization to put more effort into maximizing responses from those surveyed, perhaps via more effort invested in survey design and pre-testing, or perhaps via more detailed nonresponse follow-up. Greater survey accuracy can result when the sampling error is more than offset by a decrease in nonresponse and other biases, perhaps due to increased response rates.

What does all of this have to do with Internet-based and E-mail surveys? Before the Internet, large surveys were generally expensive to administer and hence careful thought was given to how to best conduct the survey in order to maximize information accuracy while minimizing costs. With Internet surveys, the marginal cost of sending an electronic survey to an entire e-mail list, for example, compared to only a subset of the list, can be virtually zero. Furthermore, with virtually instantaneous access to huge numbers of people on the Internet, researchers can collect large amounts of data relatively easily and cheaply. This seems to be an easy argument in favour of conducting census surveys or for simply surveying large numbers of individuals without regard to how the individuals were recruited into the sample.

This has lead some researchers, marketers, and others to default to using convenience samples or to conduct weak census surveys without giving much thought to the alternatives and the potential trade-offs they might be making in the accuracy of their survey results. But, as we have just mentioned and will examine in some detail in this chapter, a large number of survey responses is not a guarantor of accuracy.

Conducting surveys, as in all forms of data collection, requires making compromises. We can never collect all the data we would like with perfect accuracy. Hence, it is thus critical for researchers to have a firm grasp of the trade-offs they implicitly or explicitly make when choosing a sampling method for collecting their data.

DRAFT

An Overview of Sampling

There are many ways to draw samples from a population – and there are also many ways that sampling can go awry. We intuitively think of a good sample as one that is representative of the population from which the sample has been drawn. By 'representative' we do not necessarily mean that the sample matches the population in terms of observable characteristics but rather that the results from the data we collect from the sample are consistent with the results we would have gotten if we had collected data on the entire population.

Of course, the phrase 'consistent with' is vague and, if this was an exposition of the mathematics of sampling, would require precise mathematical definition. However, we will not seek to cover all of the mathematical details of survey sampling in this chapter. Rather, in this subsection we will describe the sampling process and discuss the main issues in characterizing the accuracy of a sample, with a particular focus on terminology and definitions, in order that we can put the subsequent discussion about Internet-based surveys in an appropriate context.

Readers interested in a detailed treatment of the mathematics should consult one of the classic texts such as Kish (1965) or Cochran (1977); readers interested in a summary treatment of the mathematics and/or a more detailed discussion of the sampling process may consult a number of other texts, such as Fink (2003) or Fowler (2002).

Sources of Error in Surveys

The primary purpose of a survey is to gather information about a population. However, even when a survey is conducted as a census, the results can be affected by several sources of error. A good survey design seeks to reduce all types of errors – not only the sampling error arising from surveying a sample of the population.

Survey error is commonly characterized in terms of the precision of statistical estimates. Yet charactering survey error only in terms of standard errors and response rates misses other ways in which error can enter into a

DRAFT

survey process. Table 1 below lists the four general categories of survey error as presented and defined in Groves (1989) as part of his 'Total Survey Error' approach.

Source of Error	Definition
Coverage	'...from the failure to give any chance of sample selection to some persons in the population'.
Sampling	'...from heterogeneity on the survey measure among persons in the population'.
Nonresponse	'...from the failure to collect data on all persons in the sample'.
Measurement	'...from inaccuracies in responses recorded on the survey instruments. These arise from: (a) effects of interviewers on the respondents' answers to survey questions; (b) error due to respondents, from the inability to answer questions, lack of requisite effort to obtain the correct answer, or other psychological factors; (c) error due to the weakness in the wording of survey questionnaires; and, (d) error due to effects of the mode of data collection, the use of face to face or telephone communications'.

Table 1. *Sources of survey error according to Groves (1989).*

Errors of coverage occur when some part of the population cannot be included in the sample. To be precise, Groves specifies three different populations:

1. The *population of inference* is the population that the researcher ultimately intends to draw conclusions about.
2. The *target population* is the population of inference less various groups that the researcher has chosen to disregard.
3. The *frame population* is that portion of the target population that can be enumerated prior to the selection of the sample. A broader definition proposed by Wright and Tsao (1983) is that portion of the target population which the survey materials or devices delimit, identify, and subsequently allow access to.

DRAFT

The *survey sample* then consists of those members of the sampling frame that were chosen to be surveyed and *coverage error* is the difference between the frame population and the population of inference. Figure 1 illustrates the various quantities.

The two most common approaches to reducing coverage error are:

(a) obtaining as complete a sampling frame as possible (or employing a *frameless sampling strategy* in which most or all of the target population has a positive chance of being sampled), and

(b) *post-stratifying* to weight the survey sample to match the population of inference on some observed key characteristics.

Sampling error arises when a subset of the target population is surveyed yet inference is made about the whole population. Assuming no difference between the population of inference and the target population, the sampling error is simply a quantification of the uncertainty in the sample statistic. This uncertainty can be decomposed into a variance component and a bias component. Groves (1989) states that variance characterizes the variability in the sample statistic that arises from the heterogeneity on the survey measure among the population. On the other hand, bias results from a systematic difference between the sample statistic and the actual population parameter of interest.

Nonresponse errors occur when data is not collected on either entire respondents (*unit nonresponse*) or individual survey questions (*item nonresponse*). Groves (1989) calls nonresponse 'an error of nonobservation'. The response rate, which is the ratio of the number of survey respondents to the number sampled, is often taken as a measure of goodness. Higher response rates are taken to imply a lower likelihood of *nonresponse bias*, where such a bias exists if there is a substantial difference between the observed survey results and what would have resulted if the entire sample had responded to the survey.

Measurement error arises when the survey response differs from the 'true' response. For example, respondents may not answer sensitive questions

DRAFT

honestly for a variety of reasons, or respondents may make errors in answering questions or misinterpret questions. These measurement errors may be mitigated or exacerbated by the mode of data collection.

Sampling Methods

Survey sampling can be grouped into two broad categories: *probability-based sampling* (also loosely called 'random sampling') and *non-probability sampling*. A probability-based sample is one in which the respondents are selected using some sort of probabilistic mechanism and where the probability with which every member of frame population could be selected into the sample is known. The sampling probabilities do not necessarily have to be equal for each member of the sampling frame. Often the population is enumerated in a sampling frame from which the potential respondents are randomly selected according to probabilities defined by the researcher to meet certain research criteria. This enumeration may be an actual enumeration, in the form of some type of complete list of the population, or it may be implied, such as when a multi-stage sampling scheme is used and only the members of selected primary sampling units are actually enumerated.

Types of probability samples include:

- *Simple random sampling* (SRS) is a method in which any two groups of equal size in the frame population are equally likely to be selected. Mathematically, simple random sampling selects *n* units out of a frame population of size *N* such that every sample of size *n* has an equal chance of being drawn.

  For example, given a list of 3,000 e-mail addresses of possible survey respondents, a researcher would like to survey a sample of 200. Then a simple random sample is one in which every possible group of 200 e-mail addresses out of the 3,000 are equally likely.

- *Stratified random sampling* is useful when the population is comprised of a number of homogeneous groups. In these cases, it can be either practically or statistically advantageous (or both) to first stratify the population into the homogeneous groups and then use SRS to draw samples from each group.

DRAFT

Mathematically speaking, the population of N units is first divided into k mutually exclusive groups or *strata*, N_1, N_2,…, N_k. Then samples, perhaps of different sizes, are drawn independently from each strata.

For example, perhaps the 3,000 e-mail addresses just described could be grouped into computer power users (N_1=100), average computer users (N_2=1,800), and computer novices (N_3=1,100). A sample drawn from such a population using SRS will very likely result the selection of only a few power users. (For example, a SRS of size n=200 will likely yield about six or seven power users.) If it is important to obtain a larger sub-sample of power users in the overall sample, one way to do this is to draw a SRS from each of the user strata according to pre-defined sub-sample sizes, n_1, n_2, and n_3.

- *Cluster sampling* may be applicable when the natural sampling unit is a group or cluster of individual units. For example, sometimes to survey individuals it is useful to use household as the sampling unit. The household is the cluster and in this scheme a sample of households are first selected and then individuals are randomly selected from within each household. In surveys of Internet users, it is sometimes useful or convenient to first sample by discussion groups or Internet domains and then to sample individual users.

- *Systematic Sampling* is the selection of every *k*th element from a sampling frame or every *k*th visitor or customer. When sampling from a sampling frame, *k* is chosen to be the smallest integer of the ratio of the number of elements in the population to the desired sample size. That is, *k* is the smallest integer of *N/n.*

  For example, suppose we want to sample 200 e-mail addresses from the sampling frame that contains 3,000 e-mail addresses. Then we have that 3,000/200=15, so every 15th e-mail address in the list is chosen after beginning at a random point between 1 and 15. If the random starting point is 5, then the e-mail addresses selected are 5, 20, 35, 60, …, 2,975, 2,990.

  Now, if there were 3,050 e-mail addresses, 3,050/200=15.25, so we take every 15th address. (If we were to take every 16th address, 200*16=3,200,

DRAFT

so that the last few e-mail addresses 'chosen' will not exist and the resulting sample size will be smaller than desired. That is, we would only select 190 e-mail addresses.) However, note that 200*15=3,000, which means that if we use a random starting point between 1 and 15 the last 50 e-mail address will never have a chance of being selected. Hence, the random starting point should now be chosen to be between 1 and 65 to ensure that every e-mail address has some chance of being selected.

Systematic sampling can also be applied to sampling sequential visitors, say to a website, where it has the advantage that a sampling frame does not need to be assembled beforehand. The resulting sample is considered to be a probability sample so long as the sampling interval does not coincide with a pattern in the sequence being sampled and a random starting point is chosen.

There are important analytical and practical considerations associated with how one draws and subsequently analyzes the results from each of these types of probability-bases samples that space limitations preclude covering here. Readers interested in such details should consult a text such as Kish (1965), Cochran (1977), Fink (2003), or Fowler (2002).

Non-probability samples, sometimes called *convenience samples,* occur when the probability that every unit or respondent was included in the sample cannot be determined. As the name implies, such samples are often used because it is somehow convenient to do so. A distinguishing feature between probability and non-probability samples is who decides whether an individual is included in the sample. For probability samples, the surveyor chooses and applies the probability mechanism by which the sample is selected. The individuals in the population have no control over this process.

Non-probability-based samples often require much less time and effort, and thus usually are less costly to generate, but generally do not support statistical inference. However, non-probability-based samples can be useful to researchers in other ways. For example, early in the course of research, responses from a convenience sample might be useful in developing research hypotheses. Responses from convenience samples might also be useful for

DRAFT

identifying issues, defining ranges of alternatives, or collecting other sorts of non-inferential data.

Specific types of non-probability samples include the following.

- *Judgment sampling* is a type of convenience sampling in which the researcher selects the sample based on his or her judgment. For example, a researcher may decide to draw the entire random sample from one 'representative' Internet user community, even though the population of interest includes all Internet users. This type of sampling does not allow formal statistical inference to anything other than the one user community from which the random sample was drawn, though a researcher might try to claim via his or her expert judgment that the results generalize to some broader population. Judgment sampling can also be applied in even less structured ways without the application of any random sampling, in which case statistical inference cannot be applied at all.
- *Quota sampling* is the non-probability sampling equivalent of stratified sampling. Like stratified sampling, the researcher first identifies the strata and their proportions as they are represented in the population. Then the choice of respondents is left up to the individual or individuals conducting the survey, where the interviewers are only required to match specific quotas within each strata.

  Because the choice of respondents is left up to the survey interviewers, subtle and perhaps subconscious biases may creep into the selection of the sample. In the section on Historical Survey Gaffes, we will examine a specific example of how interviewer bias in quota sampling lead the Gallup organization to incorrectly project the winner of the 1948 US presidential election.
- *Snowball sampling* is often used when the desired sample characteristic is rare and it is extremely difficult or prohibitively expensive to locate respondents by other means. Under these conditions, a random sample will be inefficient because the resulting sample is unlikely to contain many (or perhaps even any) respondents with the desired characteristics.

DRAFT

Snowball sampling relies on referrals from initial respondents to generate additional respondents. While this technique can dramatically lower search costs, it comes at the expense of introducing bias because the technique itself substantially increases the likelihood that the sample will not be representative of the population.

Bias versus Variance

If a sample is systematically not representative of the population of inference in some way then the resulting analysis can be *biased*. Bias means that the estimates resulting from the sample may not correctly measure the desired characteristic in the population. For example, results from a survey of Internet users about personal computer usage is not likely accurately quantify computer usage in the general population simply because the sample is only comprised of those who use computers.

The issue with bias is that no matter how big the sample, the estimated answer is not going to reflect the true answer if the entire population was observed. For example, the estimate of average computer usage from the Internet survey is going to overestimate the average usage in the entire US population *regardless of how many Internet users are surveyed*.

Variance, on the other hand, is simply a measure of variation in the observed data. It is a reflection of the reality that a different sample would very likely result in observing different values of the survey statistics. However, it can be used to quantify the variability of the survey statistics and hence provide a measure of the precision of the survey estimates. That is, from the variance sampling error can be calculated in terms of the standard errors of the statistics.

When thinking most simply about the precision of statistical estimates drawn via probabilistic sampling mechanisms, such estimates are improved by larger sample sizes, which can be achieved either by selecting a larger sample of potential respondents, minimizing their nonresponse via various mechanisms, or through a combination of both.

DRAFT

Randomization is used to minimize the chance of bias. The idea is that by randomly choosing potential survey respondents the sample is likely to contain a little bit of everything or everyone. For example, if we wanted to estimate the average height of some population that contained a roughly equal proportion of men and women, using randomization it would be very unlikely that the sample only contained men or only women.

The 'magic' of randomization is that the sample is very likely to 'look like' the population, even in terms of those characteristics that cannot be observed or known. The beauty of randomization is that it turns bias into variability. As we said previously, larger samples will not fix bias. However, once we turn bias into variance we can then control variance (to the extent desired and budget will allow) by increasing the sample size.

Some Important Sources of Bias

Bias can creep into survey results in many different ways. In the absence of significant nonresponse, the probabilistic mechanism is assumed to minimize the possibility of bias. Convenience sampling, on the other hand, is generally assumed to have a stronger potential to give a biased samples because the mechanism that generated the sample is not understood and hence cannot be adjusted for

However, even with randomization, surveys can encounter other kinds of bias when they are studying people. For example, interviewees may be inclined to over or understate certain things (*sensitivity bias*), particularly with socially delicate questions (such as questions about income or sexually orientation, for example). Here we just focus on some of the more common sources of bias related to sampling.

- *Frame coverage bias* occurs when the frame misses some important part of the population. For example, a mail survey based on address lists will miss the homeless. An e-mail survey using a list of e-mail addresses will miss those without an e-mail address.

DRAFT

- *Selection bias* is an error in how the individual or units are chosen to participate in the survey. It can occur in convenience samples, for example, when those with strong opinions choose to participate in a survey out of proportion to their numbers in the general population. It can also occur if survey participation depends on the respondents having access to particular equipment. For example, Internet-based surveys will miss those without Internet access.
- *Nonresponse bias* occurs if those that refuse to answer the survey are somehow systematically different that those that do answer the survey. For example, in telephone surveys, those with higher incomes often have answering machines to screen calls and are more likely not to be at home. Hence, the resulting survey responses are likely to be biased towards those with lower incomes.
- *Size bias* occurs when some units have a greater chance of being selected than others. For example, in a systematic sample of website visitors frequent site visitors are more likely to get selected into the sample than those that do not.

Historical Survey Gaffes

A famous example of a survey that reached exactly the wrong inferential conclusion as a result of bias, in this case frame coverage and nonresponse bias, is the *Literary Digest* poll in the 1936 presidential election. As described in Squires (1988), for their survey *Literary Digest* assembled a sampling frame from telephone numbers and automobile registration lists. While using telephone numbers today might result in a fairly representative sample of the population, in 1936 only one in four households had a telephone. Those households that did were the more well-to-do. Compounding this was the addition of automobile registration lists to the sampling frame which only further skewed the frame towards people with higher incomes.

Literary Digest mailed 10 million straw vote ballots, of which 2.3 million were returned, an impressively large number, but it represented less than a 25

DRAFT

per cent response rate. Based on the poll data, *Literary Digest* predicted that Alfred Landon would beat Franklin Roosevelt 55 per cent to 41 per cent. In fact, Roosevelt beat Landon by 61 per cent to 37 per cent. This was the largest error ever made by a major poll and is considered to be one of the major causes of the demise of *Literary Digest* in 1938.

In an analysis of data subsequently collected by George Gallup after the election, Squires (1988) concluded that nonresponse bias an important source of error: 'Among those who responded…a slight majority favoured Landon.' He goes on to say, 'more importantly, the initial sample was flawed; when compounded with the response bias it produced the wildly erroneous forecast of the vote percentages.'

Gallup, however, called the 1936 presidential election correctly even though he used significantly less data. But even Gallup, a pioneer in modern survey methods, didn't always get it right. The famous photograph of Harry S. Truman (Figure 1) holding up a newspaper with the headline 'Dewey Defeats Truman' is the result of survey errors in predicting the 1948 US presidential election.

In this case, Gallup used a quota sampling method in which each pollster was given a set of quotas of types of people to interview based on demographics. For example, one interviewer is told to interview 13 people and, of these, 7 should be men and 6 should be women; of the men, get 3 under 40 and 4 over 40, etc. The idea is to achieve a 'representative' sample by explicitly matching the distribution of observable characteristics in the sample to the distribution of those same characteristics in the population.

DRAFT

Figure 1. *Errors in the 1948 US presidential election predictions resulted from an improper application of quota sampling. Here President Truman holds up a copy of the Chicago Daily Tribune that mistakenly announced that Thomas Dewey had won the election.* (© Associated Press. Photograph taken by Byron Rollins on November 4, 1948 in Saint Louis, MO.)

That seems like a good idea. So, what went wrong? Among other surveying mistakes (such as the pollsters—including Roper and Crossley—stopped surveying two weeks before the election), under the quota sampling system the survey interviewers, for whatever reason, chose to interview republicans more often than democrats. This could have resulted because republicans were easier to locate or perhaps because, consciously or subconsciously, the pollsters simply preferred to interview them. Gallup predicted a Dewey win of 49.5 per cent to 44.5 per cent, but almost the opposite occurred with Truman beating Dewey with 49.5 per cent of the popular vote to Dewey's 45.1 per cent (a difference of almost 2.2 million votes).

These were not the only errors made in the 1936 and 1948 presidential election polls (for more detail, see Zetterberg (2004) or Ross (1977)), but they were significant errors, and are highlighted here to illustrate that the various biases that can creep into survey and data collection can be both subtle and non-trivial.

Sampling Methods for Internet-based Surveys

In this section we now focus in on the specific types of Internet-based surveys and the sampling methods that are applicable to each. Here we mainly focus on describing the types of surveys and sampling methods, only briefly mentioning some of the issues associated with actual implementation. In the section that follows, Issues and Challenges in Internet-based Survey Sampling we discuss these matters in more detail.

With any survey, it is important to begin with a careful assessment of the purpose of the survey. It is only in the context of the survey purpose that a judgment can be made about whether a particular type of survey or sampling method is appropriate. Couper (2000) perhaps said it best,

> *Any critique of a particular Web survey approach must be done in the context of its intended purpose and the claims it makes. Glorifying or condemning an entire approach to survey data collection should not be done on the basis of a single implementation, nor should all Web surveys be treated as equal.*

In this section, we will largely concentrate on differentiating whether particular sampling methods and their associated surveys allow for generalization to populations of inference or not. We will provide examples of some surveys that were done appropriately and well and others that were less so. Examples that fall into the latter category should not be taken as a condemnation of a particular survey or sampling method, but rather as illustrations of inappropriate application, execution, analysis, etc. Furthermore, simply because a particular method does not allow for generalizing beyond the sample does not imply that the methods and resulting data are not useful in other ways for answering other types of research questions.

DRAFT

Types of Internet-based Surveys

Similar to Couper (2000), we classify the most common types of Internet-based surveys in Table 2 according to whether they are based on probability or non-probability sampling methods and which Internet-based survey mode or modes apply or are most generally used. For example, while it is feasible to conduct an entertainment poll by e-mail, virtually all such polls are conducted via web-based surveys.

Sampling Method	Web	E-mail
Probability-based		
Surveys using a list-based sampling frame	✓	✓
Surveys using non-list-based random sampling	✓	✓
Intercept (pop-up) surveys	✓	
Mixed mode surveys with Internet-based option	✓	✓
Pre-recruited panel surveys	✓	✓
Non-probability		
Entertainment polls	✓	
Unrestricted self-selected surveys	✓	
Surveys using 'Harvested' E-mail Lists (and Data)	✓	✓
Surveys using volunteer (opt-in) panels	✓	

Table 2. *Types of Internet-based surveys and associated sampling methods*.

As we discussed in the previous section, one of the major distinguishing features between probability and non-probability samples is whether it is left up to the individual to choose to participate in the survey ('opt in') or not. For probability samples, the respondent is not given this option. He or she is selected via some probabilistic method established by the survey researcher. For non-probability based methods, either a convenience sample is drawn, from which the probabilities that various members of the population being included in the sample cannot be determined, or the survey is distributed or advertised in some manner and it is left up to those exposed to the survey to choose to opt in.

While in a probability-based survey participants can choose not to participate in the survey ('opt out'), in non-probability based surveys respondents

DRAFT

must frequently decide to opt-in. In both cases, it is possible to have nonresponse bias, but in non-probability surveys it has the potential to be much greater since it is likely that those that opt-in are not representative of the general population (they are likely to have stronger views about the subject being surveyed, for example). Furthermore, in non-probability surveys there is often no way to assess the potential magnitude of the nonresponse bias since there is generally no information on those that chose *not* to opt-in.

Surveys Using a List-based Sampling Frame

Sampling for Internet-based surveys using a list-based sampling frame can be conducted just like one would for a traditional survey using a sampling frame. Simple random sampling in this situation is straightforward to implement and requires nothing more than contact information (generally an e-mail address for an Internet-based survey) on each unit in the sampling frame. Of course, though only contact information is required to field the survey, having as much additional demographic and other information about each unit in the sampling frame is desirable to assess (and perhaps adjust for) nonresponse effects.

While Internet-based surveys using list-based sampling frames can be conducted either via the web or by e-mail, if an all-electronic approach is preferred, the invitation to take the survey will almost always be made via e-mail. And, because e-mail lists of general populations are simply not available, this survey approach is most applicable to large homogeneous groups for which a sampling frame with e-mail addresses can be assembled (for example, universities, government organizations, large corporations, etc.). Coupler (2000) calls these 'list based samples of high-coverage populations' and the population of inference in these situations may be limited.

To illustrate, Fahmy et al. (2005) report the results of an all-electronic survey on the ethics of modifying digital photographs by the media. Here the authors used a sampling frame which contained the e-mail addresses for the American Society of Magazine Editors and invited *all* of the society's members to complete the survey. (This is an example of a sample in which the probability of

DRAFT

being selected into the sample is one for all members of the sampling frame.) Personalized e-mails were sent to all 822 members inviting them to take a web-based survey and up to three follow-up e-mails were sent to nonrespondents. In all, the authors were able to contact 93 per cent of the society's members (765) and received survey responses from 210, for a response rate of just over 27 per cent.

Of course, the response rate raises questions about nonresponse bias, but the point here is simply to illustrate the application of a list-based sample for an Internet-based survey. With Internet-based surveys, because it is just as easy to send e-mail to the entire list as it is to a portion of a list, the result is an implicit or explicit preference for nonresponse bias over sampling error. That researchers choose to attempt censuses rather than sample – a seemingly common practice in Internet-based surveys when an e-mail address list is available – is a point we will return to later.

In more complicated sampling schemes, such as a stratified sampling, auxiliary information about each unit, such as membership in the relevant strata, must be available and linked to the unit's contact information. And more complicated multi-stage and cluster sampling schemes can be difficult to impossible to implement for Internet-based surveys. First, to implement without having to contact respondents will likely require significant auxiliary data available as part of the sampling frame which is unlikely to be available except in the case of specialized populations. Second, if contact is required, then the researchers are likely to have to resort to the telephone or mail in order to ensure sufficient coverage and response rates are achieved.

An example of multi-stage sampling procedure used for an online survey of real estate journalists for which no sampling frame (initially) existed is reported by Jackob et al. (2005). For this study, the researchers first assembled a list of publications what would have journalists relevant to the study. From this list a stratified random sample of publications was drawn, separately for five European countries. They then contacted the managing editor at each sampled publication and obtained the necessary contact information on all of the journalists that were

DRAFT

'occupied with real estate issues'. All of the journalists identified by the managing editors were then solicited to participate in a web survey. Jackob et al. (2005) concluded 'that conducting an international comparative online survey takes a lot of effort especially during the phase of preparation and planning'.

Surveys Using Non-list-based Random Sampling

Non-list-based random sampling methods allow for the selection of a probability-based sample without the need to actually enumerate a sampling frame. With traditional surveys, random digit dialing (RDD) is a non-list-based random sampling method that is used mainly for telephone surveys.

There is no equivalent of RDD for Internet-based surveys. For example, it is not possible to generate random e-mail addresses (see the Issues and Challenges in Internet-based Survey Sampling section). Hence, with the exception of intercept surveys, Internet-based surveys requiring non-list-based random sampling depend on contacting potential respondents via some traditional means such as RDD.

However, contacting respondents via a traditional media introduces other complications and costs. For example, surveyors must either screen potential respondents to ensure they have Internet access or field a survey with multiple response modes. Surveys with multiple response modes introduce further complications, both in terms of fielding complexity and possible mode effects (see the Issues and Challenges in Internet-based Survey Sampling section).

Intercept Surveys

Intercept surveys on the web are pop-up surveys that frequently use systematic sampling for every *n*th visitor to a website or web page. These surveys seem to be most relevant as customer satisfaction surveys or marketing surveys. This type of systematic sampling can provide information that is generalizable to particular populations, such as those that visit a particular website/page. Furthermore, the surveys can be restricted to only those with certain IP (Internet protocol) addresses, allowing one to target more specific

DRAFT

populations of inference than just all visitors to a website/page, and 'cookies' can be used to restrict the submission of multiple surveys from the same computer.

A potential issue with this type of survey is nonresponse. Coomly (2000) reports typical response rates in the 15 to 30 per cent range with the lowest response rates occurring for poorly targeted and/or poorly designed surveys. The highest response rates were obtained for surveys that where relevant to the individual, either in terms of the particular survey questions or, in the case of marketing surveys, the commercial brand being surveyed.

As discussed in Couper (2000), an important issue is that with intercept surveys there is no way to assess nonresponse bias, simply because no information is available on those that choose not to complete a survey. Coomly (2000) hypothesizes that responses may be biased towards those who are more satisfied with a particular product, brand, or website; towards those web browsers who are more computer and Internet savvy; and, away from heavy Internet browsers who are conditioned to ignore pop-ups. Another source of nonresponse bias for intercept surveys implemented as pop-up browser windows may be pop-up blocker software, at least to the extent that pop-up blocker software is used differentially by various portions of the web browsing community.

Pre-recruited Panel Surveys

Pre-recruited panel surveys are, generally speaking, groups of individuals who have agreed in advance (and hence are 'pre-recruited') to participate in a series of surveys. For Internet-based surveys requiring probability samples, these individuals must be pre-recruited via some means other than the web or e-mail – most often it is by telephone or postal mail.

For an effort that is a longitudinal series of surveys, researchers may recruit panel members specifically for that effort. For smaller efforts or for single surveys, there are a number of companies that maintain panels of individuals, pre-recruited via a probabilistic sampling methodology, from which sub-samples can be drawn according to a researcher's specification. Knowledge Networks,

DRAFT

for example, recruits all of its panel members via telephone using random digit dialing, and it provides equipment and Internet access to those that do not have it in an attempt to maintain a panel that is a statistically valid cross-section of the population. (See Pineau and Dennis, 2004, for additional detail.)

Pre-recruited panel surveys that use probabilistic sampling, such as Knowledge Networks, can be an attractive option to researchers who desire to field and Internet-based survey but that require a sample that can be generalized to populations outside of the Internet-user community. Pre-recruited panels that have been Internet-enabled can combine the speed of Internet-based surveys while simultaneously eliminating the often-lengthy recruitment process that would normally be required.

However, pre-recruited panels are not without their potential drawbacks. In particular, researchers should be aware that long-term panel participants may respond differently to surveys and survey questions (called 'panel conditioning' or 'time-in-sample bias') than first time participants for a variety of reasons. Also, nonresponse can be an issue if the combined loss of potential respondents throughout all of the recruitment and participation stages is significant. However, as Couper (2000) concludes, '… in theory at least, this approach begins with a probability sample of the full (telephone) population, and assuming no nonresponse error permits inference to the population...'

Entertainment Polls

Internet-based entertainment polls are 'surveys' conducted purely for their entertainment value (though they are sometimes passed off to be more than what they are). In the Internet-based realm, they largely consist of websites where site visitors can respond to one or more surveys that are posted and which anyone can complete without restriction. The telephone equivalent of these types of polls is the 900 number call-in poll such as those advertised on various television shows where viewers can vote for their favourite contestant or character. Examples of Internet-based entertainment polls include The Weekly Web Poll (www.weeklywebpoll.com) and, perhaps disappointingly, almost every

DRAFT

local news channel's website. Of course, Internet-based entertainment polls are no more scientific than 900 number call-in telephone polls.

Surveys using 'Harvested' E-mail Lists

Harvested e-mail lists are those collected by some means from the web, either automatically or manually for e-mail addresses that are posted on websites, and also wittingly or unwittingly solicited from individuals browsing the web. There are many commercial entities ('e-mail brokers') that sell lists of e-mail addresses or access to lists or e-mail addresses (just Google 'buy e-mail list').

Researchers can also assemble various lists of e-mail addresses directly from the web. For example, lists of Yahoo e-mail address holders by name or geographic area can be created by anyone via the Yahoo! People Search (http://email.people.yahoo.com/py/). Similarly, World Email.com (www.worldemail.com) has an e-mail search feature by name. However, it is important to note that harvesting e-mail addresses and distributing unsolicited e-mail related to surveys violates the ethical standards of the Council of American Survey Research Organizations (CASRO). See Figure 2 below or the complete CASRO Code of Standards and Ethics for Survey Research at www.casro.org/codeofstandards.cfm.

Samples derived from harvested e-mail lists are non-probability samples because they are based on a convenience sample of e-mail addresses which are often an aggregation of addresses collected in various ways. For example, Email Marketing Blitz (www.email-marketing-blitz.com/customized_email_list.htm) says, 'Our targeted optin [sic] email lists are updated monthly and gathered through special interest websites, entertainment websites and special alliances'. Such samples should not be confused with list-based probability samples where the e-mail addresses in the list-based sample represent a (virtually) complete list of the e-mail addresses of some population.

The efficacy of sending unsolicited surveys to a list of purchased or otherwise procured list e-mail addresses is questionable. Not only do e-mail

DRAFT

addresses turn over quite frequently, but many of those on the list were likely recruited either without their knowledge or they may have inadvertently agreed by failing to uncheck a box when they signed up for something else. In any case, response rates are likely to be extremely low. In addition, unsolicited surveys are essentially another form of spam and may either be unethical or even illegal in some parts of the world.

However, Lockett and Blackman (2004) demonstrate that it is not always necessary to harvest e-mail addresses to collect data via the web. They present a case study of Xenon Laboratories, an Internet-based financial services firm that employed a novel approach to market research. In particular, Xenon Laboratories wanted to collect data on foreign exchange charges by credit card companies on business travellers. Xenon Laboratories recognized that neither the travellers nor the credit card companies were likely to respond to a survey on this topic, whether fielded over the web or otherwise.

Instead Xenon Laboratories developed the Travel Expenses Calculator (www.xe.com/tec) and the Credit Card Charges Calculator (www.xe.com/ccc) and posted them on the web for anyone to use for free. These tools help foreign business travellers to accurately calculate the cost of a business expense receipt in terms of their own currency. Lockett and Blackman (2004) say, 'On the basis of this information [input by those using the calculators] it is possible to conduct basic market research by aggregating the inputted calculations. Xenon is now in the unique position to analyse whether or not the different card providers employ the same charging levels and whether or not these companies' charge structures vary according to geographical region'. They go on to conclude, 'This value-added approach, which is mutually beneficial to both parties, is an important and novel approach to market research'.

Unrestricted Self-selected Surveys

Unrestricted self-selected surveys are surveys that are open to the public for anyone to participate in. They may just be posted on a website so that, anyone browsing through that page may choose to take the survey, or they may

DRAFT

be promoted via website banners or other Internet-based advertisements, or they may be publicized in traditional print and broadcast media. Regardless of how they are advertised (or not), the key characteristics of these types of surveys are that there are no restrictions on who can participate and it is up to the individual to choose to participate (opt-in).

For example, Berson et al. (2002) conducted a web-based survey 'to better understand the risks to adolescent girls online' by posting a link to their survey on the *Seventeen Magazine Online* website. Via the survey, the authors collected data on 10,800 respondents with 'identified behaviours that put them at risk'. The researchers were careful to appropriately qualify their results:

> *The results highlighted in this paper are intended to explore the relevant issues and lay the groundwork for future research on youth in cyberspace. This is considered an exploratory study which introduces the issues and will need to be supplemented with ongoing research on specific characteristics of risk and prevention intervention. Furthermore, the generalizability of the study results to the larger population of adolescent girls needs to be considered. Due to anonymity of the respondents, one of the limitations of the research design is the possibility that the survey respondents did not represent the experience of all adolescent girls or that the responses were exaggerated or misrepresented.*

As with entertainment polls, unrestricted, self-selected surveys are based on a form of convenience sampling. As such, the results cannot be generalized to a larger population, but as Berson, Berson, and Ferron illustrate, that does not necessarily negate their usefulness for research.

For example, the web can facilitate access to individuals that are difficult to reach either because they are hard to identify, locate, or perhaps exist in such small numbers that probability-based sampling would be unlikely to reach them in sufficient numbers. Coomber (1997) describes such a use of the web, using it to field a survey to collect information from drug dealers about drug adulteration/dilution. By posting invitations to participate in a survey on various drug-related discussion groups, Coomber collected data from 80 survey respondents (that he deemed reliable) located in 14 countries on four different continents. The sample is certainly not generalizable, but it also provided a data

DRAFT

that is unlikely to be collected any other way, and which Coomber found consistent with other research.

In addition, Alvarez et al. (2002) propose that these types of non-probability samples can be useful and appropriate for conducting experiments (say, in the design of web pages or web surveys) by randomly assigning members of the sample to control and experimental groups. In terms of psychology experiments, Siah (2005) states, 'For experimental research on the Internet, the advantage of yielding a heterogeneous sample seems persuasive considering that the most common criticism on psychological research is its over-reliance on college student samples'. But he goes on to say, 'Nonetheless, it is worthwhile to note that the argument of conducting internet-based experiments for the benefit of enhancing generalizability misses the target in experimental research, given that this quality adds little to the rigor of experimental studies'.

Volunteer (Opt-in) Panels

Volunteer (opt-in) panels are similar in concept to the pre-recruited panels, except in volunteer panels the participants are not recruited. Rather, they choose to participate, perhaps after coming across a solicitation on a website. In this regard, volunteer panels are similar to entertainment polls except that those that opt-in do so to take a continuing series of surveys. Harris Interactive manages such a volunteer panel. On its website (www.harrispollonline.com/became.asp) they state, 'You may have become a member of the Harris Poll Online in one of several ways:

- By registering directly with us through our website (http://www.harrispollonline.com); or
- By opting in to participate in the Harris Poll Online as a result of an offering made in conjunction with one of our many online partners'.

Often these panels are focused on market research, soliciting consumer opinions about commercial products, and participants sometimes do it for monetary incentives. For example, the Harris Online Poll states,

DRAFT

We offer the opportunity to earn HIPoints for the majority of our studies. On occasion a study will be conducted that will not have HIPoints associated with it, but this only occurs in exceptions. Once you've accumulated enough points you may redeem them for your choice of a variety of great rewards (www.harrispollonline.com/benefit.asp).

Issues and Challenges in Internet-based Survey Sampling

All survey modes have their strengths and weaknesses; Internet-based surveys and e-mail surveys are no different in this regard. The various strengths and weaknesses are more or less important depending on the survey's purpose.

For example, a frequently cited weakness of Internet-based surveys is that there is no sampling frame for the general population and that they are subject to significant coverage errors for those portions of the population without Internet access. While this may indeed be a major issue for a survey that intends to generalize to an entire country's population, it is much less of an issue or not an issue at all for other types of populations (for example, a corporate employee survey in a firm where every employee has a company computer and the company has a centralized e-mail address directory).

Drawing an appropriate sample that will provide the necessary and appropriate data for answering the relevant research objective or marketing question is critical. Hence, in this section we will focus on the strengths and weaknesses that are related to the issues and challenges related to sampling for Internet-based surveys.

Sampling Frame and Coverage Challenges

As we just discussed, a frequent impediment for conducting large-scale Internet-based surveys is the lack of a sampling frame. Simply put, no single registry or list of e-mail addresses exist and thus list-based sampling frames are generally only available for specific populations (government organizations, corporations, etc.).

Compounding this difficulty, and leaving aside the issue of population coverage to be discussed shortly, it is impossible to employ a frameless sampling

DRAFT

strategy since for all practical purposes one cannot assemble random e-mail addresses. Of course, it is theoretically possible to 'construct' e-mail addresses by repeatedly randomly concatenating letters, numbers, and symbols, but the sheer variety of e-mail addresses means most of the constructed addresses will not work. More importantly, the unstructured nature of the Internet means that even if one could tolerate the multitude of undeliverable e-mail messages that would result, they could not be useful as the basis for a probability sample.

In terms of coverage, it is widely recognized that Internet-based surveys using only samples of Internet users do not generalize to the general public. While Internet penetration into households continues at a rapid pace, the penetration is far from complete (compared to, say, the telephone) and varies widely by country and region of the world. The point is, if the target of inference is the general public, considerable coverage error remains for any sample drawn strictly from Internet users.

Now, even if there is minimal coverage error for a particular Internet-based survey effort, when using only an Internet-based survey mode it is important to ensure *all* of the target population is sufficiently computer literate and has both regular and easy access to the Internet to facilitate responding to the survey. Simply put, just because an organization maintains a list of e-mail addresses for everyone in the organization it does not necessarily follow that every individual on the list has equal access. Lack of equal access could result in significant survey selection and nonresponse biases.

For some surveys it may be (fiscally and operationally) possible to contact respondents by some mode other than e-mail, such as mail or the telephone. In these cases the survey target population can be broader than that for which an e-mail sampling frame is available, up to and including the general population. But at present such a survey must also use multiple survey modes to allow respondents without Internet access the ability to participate. Mixed-mode surveys may also be useful for alleviating selection bias for populations with uneven or unequal Internet access.

DRAFT

However, mixed-mode surveys are subject to other issues. Two of the most important are mode effects and respondent mode preferences. Mode effects arise when the type of survey affects how respondents answer questions. Comparisons between Internet-based surveys and traditional surveys have found conflicting results, with some researchers reporting mode effects and others not. See, for example, the discussion and results in Schonlau et al. (2004: 130). Though not strictly a sampling issue, the point is that researchers should be prepared for the existence of mode effects in a mixed-mode survey. The next chapter will explore the issues of combining data from Internet-based and traditional surveys in greater detail.

In addition, when Internet-based surveys are part of a mixed-mode approach, it is important to be aware that the literature currently seems to show that respondents will tend to favour the traditional survey mode over an Internet-based mode. See, for example, the discussions in Schonlau et al. (2002) and Couper (2000: 486-487). Fricker and Schonlau (2002) in a study of the literature on web-based surveys found 'that for most of the studies respondents currently tend to choose mail when given a choice between web and mail. In fact, even when respondents are contracted electronically it is not axiomatic that they will prefer to respond electronically'.

The tendency to favour non-Internet-based survey modes lead Schonlau et al. (2002: 75) to recommend for mixed mode mail and web surveys that:

> *…the most effective use of the Web at the moment seems to involve a sequential fielding scheme in which respondents are first encouraged to complete the survey via the Web and then nonrespondents are subsequently sent a paper survey in the mail. This approach has the advantage of maximizing the potential for cost savings from using Internet while maintaining the population coverage and response rates of a mail survey.*

Web-based Recruitment Issues and Effects

Whether e-mail addresses are constructed, assembled from third party sources, or harvested directly from the web, there is the issue of unsolicited surveys as spam. For example, Sheehan (1999) conducted a survey with e-mail

DRAFT

addresses harvested from www.Four11.com and stated, 'Several individuals receiving the solicitation e-mail censured the researchers for sending out unsolicited e-mails, and accused the researchers of "spamming"'. They further recounted that 'One [ISP] system operator [who observed a large number of e-mail messages originating from a single address] then contacted his counterpart at our university'.

Recruiting survey participants using unsolicited e-mail is at a minimum a potential ethical issue and in some countries may be illegal. Setting the question of legality aside, harvesting e-mail addresses and distributing unsolicited e-mail related to surveys is considered unethical according to the Council of American Survey Research Organizations (CASRO) Code of Standards and Ethics for Survey Research, of which an excerpt is reproduced in Figure 2 (highlighting added for emphasis).

---

3. Internet Research

a. The unique characteristics of internet research require specific notice that the principle of respondent privacy applies to this new technology and data collection methodology. The general principle of this section of the Code is that survey research organizations will not use unsolicited emails to recruit respondents for surveys.

(1) Research organizations are required to verify that individuals contacted for research by email have a reasonable expectation that they will receive email contact for research. Such agreement can be assumed when ALL of the following conditions exist.

a. A substantive pre-existing relationship exists between the individuals contacted and the research organization, the client or the list owners contracting the research (the latter being so identified);
b. Individuals have a reasonable expectation, based on the pre-existing relationship, that they may be contacted for research;
c. Individuals are offered the choice to be removed from future email contact in each invitation; and,
d. The invitation list excludes all individuals who have previously taken the appropriate and timely steps to request the list owner to remove them.

(2) Research organizations are prohibited from using any subterfuge in obtaining email addresses of potential respondents, such as collecting email addresses from public domains, using technologies or techniques to collect email addresses without individuals' awareness, and collecting email addresses under the guise of some other activity.

Figure 2. *Excerpt from Council of American Survey Research Organizations Code of Standards and Ethics for Survey Research, accessed online at www.casro.org/codeofstandards.cfm on September 30, 2006. Highlighting added.*

For a more detailed discussion of the ethical considerations and implications, see Section II of this handbook and Krishnamurthy (2002) and the references contained therein.

DRAFT

Regardless of the ethical issues, distributing an unsolicited Internet-based survey is also not without its perils. For example, Andrews et al. (2002) report on a study of 'hard-to-involve Internet users': those that lurk in but do not participate publicly in online discussion forums. In their study, an invitation to participate in a web survey was posted as a message to 375 online community discussion boards. While they collected 1,188 valid responses (out of 77,582 discussion board members), they also 'received unsolicited email offers, some of which were pornographic in content or aggressive in tone' and they had their web server hacked twice, once with the infection of a virus.

In spite of the challenges and possible perils, it is possible to recruit survey participants from the web. For example, Alvarez et al. (2002) conducted two Internet-based recruitment efforts – one using banner advertisements on web pages and another using a subscription check box. In brief, their results were as follows.

- In the first recruitment effort, they ran four 'banner' campaigns in 2000 with the intention of recruiting survey participants using web page banner advertisements. In the first campaign, which is representative of the other three, an animated banner advertisement resulted in more than 3.5 million 'impressions' (the number of times the banner was displayed), which resulted in the banner being clicked 10,652 times, or a rate of 3 clicks per 1,000 displays. From these 10,652 clicks, 599 survey participants were recruited.
- In the second recruitment effort, the authors ran a 'subscription' campaign in 2001 in which they arranged with a commercial organization to have a check box added to subscription forms on various websites. Essentially, Internet users who were registering for some service were given an opportunity to check a box on the services subscription form indicating their willingness to participate in a survey. As part of this effort, the authors conducted two recruitment drives, each of which was intended to net 10,000 subscriptions. Across the two campaigns, 6,789 new survey participants were obtained from 21,378 subscribers.

DRAFT

The good news from the Alvarez et al. (2002) study is that, even though the banner approach yielded fewer new survey participants, both methods resulted in a significant number of potential survey respondents over a relatively short period of time: 3,431 new subjects over the course of six or seven weeks from the banner campaigns and 6,789 new subjects over the course of three weeks from the subscription campaigns. Each banner subject cost about $7.29 to recruit while the subscription subjects cost only $1.27 per subject. (Unfortunately, the authors did not present any data on survey completion rates, so we do not know whether there differences between the two samples that might favour one over the other.)

The bad news, is that the two groups differed significantly in all of the demographic categories collected (gender, age, race, and education) and they differed in how they answered questions on exactly the same survey. In addition, both groups differed significantly from the demographics of the Internet population as measured by the August 2000 Current Population Survey. The problem, of course, is that there are clear effects associated with how subjects are recruited, such that the resulting samples are different even from the general Internet population. Shillewaert et al. (1998) found similar recruitment method biases. Hence, while it is possible to ethically recruit survey participants from the web, it seems that the recruitment methodology affects the types of individuals that self-select into the sample.

Improving Response Rates for Internet-based Surveys

Response rates have a direct effect on sampling: the higher the response rate the fewer people need to be sampled to achieve a desired number of survey completions. In addition, higher response rates are associated with lower nonresponse bias.

Unfortunately, in a summary of the academic survey-related literature up through 2001, Fricker and Schonlau (2002) concluded that ‘Web-only research surveys have currently only achieved fairly modest response rates, at least as

DRAFT

documented in the literature'. S. Fricker et al. (2005) similarly summarized the state of affairs as 'Web surveys generally report fairly low response rates'.

A good illustration of this is the Couper et al. (1999) study in which employees of five US federal government statistical agencies were randomly given a mail or e-mail survey. Comparable procedures were used for both modes, yet higher response rates were obtained for mail (68-76 per cent) than for e-mail (37-63 per cent) across all of the agencies.

Incentives are a common and effective means for increasing response rates in traditional surveys. Goritz (2006) is an excellent review of the use of incentives in survey research in which he distinguishes their use in traditional surveys from Internet-based surveys and provides a nice discussion of the issues associated with using incentives in Internet-based surveys. Open issues include:

- How best to deliver an incentive electronically.
- Whether it is better to provide the incentive in prior to a respondent taking the survey or after.
- Whether incentives have different effects for individuals taking a survey one time versus pre-recruited panel members that take a series of surveys.

Individual studies of Internet-based surveys have generally found incentives to have little or no effect. For example, Coomly (2000) found that incentives had little effect on response rates for pop-up surveys and Kypri and Gallagher (2003) found no effect in a web-based survey.

However, Göritz (2006) conducted a meta-analysis of 32 experiments evaluating the impact of incentives on survey 'response' (the fraction of those solicited to take the survey that actually called up the first page of the survey) and 26 experiments evaluating the effect of incentives on survey 'retention' (the fraction of those who viewed the first page that actually completed the survey). From the meta analysis, Görtiz concluded that 'material incentives promote response and retention in Web surveys' where 'material incentives increase the

DRAFT

odds of a person responding by 19% over the odds without incentives' and 'an incentive increased retention by 4.2% on average'.

To illustrate these effects in a concrete example, Görtiz assumes a sample of 10,000 people who, without a survey incentive, have a 40 per cent response rate and a 65 per cent retention rate. This means 2,600 people complete the survey (10,000 x 0.40 x 0.65). However, with an incentive 3,107 people can be expected to compete the survey, an almost 20 per cent improvement.

In addition to incentives, Dillman (2000) and Dillman et al. (1999) have put forward a number of survey procedural recommendations to increase survey response rates, based on equivalent methods for traditional surveys, which we will not re-cover here since they are mainly related to survey design and fielding procedures. While we do note that the recommendations seem sensible, Couper (2000) cautions that 'there is at present little experimental literature on what works and what does not'.

Bigger Samples are not Always Better

With Internet-based surveys with a list-based sampling frame, rather than sending the survey out to a sample researchers often simply send the survey out to the entire sampling frame. That is, it seems to frequently be the case that researchers conducting (all electronic) Internet-based surveys – where the marginal costs for additional surveys can be virtually nil – often fail to recognize the 'trade-off between easy, low cost access to large numbers of patients [participants] and the representativeness in the population being studied' (Soetikno et al., 1997). Simply put, for both probability and non-probability-based samples, larger sample sizes do not necessarily mean the sample is more representative of any greater population. A sample can be biased whether it is large or small.

One might argue that in these situations the researchers are attempting to conduct a census, but in practice they are foregoing a probability sample in favour of a convenience sample by allowing members of the sampling frame to

DRAFT

opt into the survey. Dillman et al. (1999) summarized this practice as follows: '…the ease of collecting hundreds, thousands, or even tens of thousands of responses to web questionnaires at virtually no cost, except for constructing and posting, appears to be encouraging a singular emphasis on the reduction of survey error'. By this Dillman et al. mean that researchers who only focus on reducing sampling error by trying to collect as large a sample as possible miss the point that it is equally important to reduce coverage, measurement, and nonresponse error in order to be able to accurately generalize from the sample data. A myopic focus on large sample sizes – and the idea that large samples equate to sample representativeness which equates to generalizability – occurs with convenience sample-based web and e-mail surveys as well.

An example of this type of focus is illustrated with'Survey2000', of a large-scale unrestricted self-selected survey conducted as a collaborative effort between the National Geographic Society (NGS) and some academic researchers. Fielded in 1998, the survey was posted on the National Geographic Society's website and participants were solicited both with a link on the NGS homepage and via advertisement in NGS periodicals, other magazines, and newspapers.

Upon completion of the effort, Witte et al. (2000) report that more than 80,000 surveys were initiated and slightly more than 50,000 were completed. While this is a relatively large number of survey completions compared to most survey efforts, it is important to remember that it is the result of an unrestricted self-selected sampling strategy and hence is still just a convenience sample that is not generalizable to any larger population. Yet, Witte et al. (2000) go to extraordinary lengths to rationalize that their results are somehow generalizeable while simultaneously demonstrating that the results of the survey generally do not correspond to known population quantities.

A related and significant concern with non-probability-based sampling methods, both for Internet-based and traditional surveys, is that survey accuracy is characterized only in terms of sampling error and without regard to the potential biases that may be present in the results. While this has always been a

DRAFT

concern with all surveys, the ease and spread of Internet-based surveys seems to have exacerbated the practice. For example, the results of an E-Poll were explained as follows:

> *THE OTHER HALF / E-Poll® Survey of 1,007 respondents was conducted January 16-20, 2003. A representative group of adults 18+ were randomly selected from the E-Poll online panel. At a 95% confidence level, a sample error of +/- 3% is assumed for statistics based on the total sample of 1,007 respondents. Statistics based on sub-samples of the respondents are more sensitive to sampling error*. (From a press release posted on the E-Poll website.)

No mention was made in the press release that the 'E-Poll online panel' consists of individuals who had chosen to participate in on-line polls nor that they were unlikely to be representative of the general population. Rather, it leaves readers with an incorrect impression that the results apply to the general population when, in fact, the margin of error for this particular survey is only valid for adult members that particular E-Poll online panel.

In Summary

A useful way to summarize which sampling methods apply to which types of Internet-based surveys is to group them by the respondent 'contact mode'. That is, every survey effort can be classified according to how the respondents are contacted (the contact mode), how they are asked to complete the survey (the response mode), and then how subsequent communication is conducted (the follow-up mode). Each of these can be executed in a different media, where the media are telephone, mail, web, e-mail, and so forth. For example, respondents may be contacted by telephone to participate in a web survey with follow-up done by mail.

This distinction is often not particularly relevant for traditional surveys since respondents that have been asked to take, say, a telephone survey have generally been contacted via the same mode. This is often the case for mail and face-to-face surveys as well. While not a strict rule – for example, a telephone survey may be preceded by mailed invitation to each survey respondent – it is

DRAFT

often the case. In comparison, given the challenges and frequent impossibility of generating random e-mail addresses in a fashion similar to RDD, as well as other challenges associated with Internet-based surveys that we have discussed in this chapter, the mode by which individuals are contacted to take a web survey is more likely to be different than the response mode.

In terms of sampling for Internet-based and e-mail surveys, what is relevant is that the sampling methodology is generally driven by the contact mode, *not* the response mode. Hence, as shown in Table 3, we can organize sampling strategies for Internet-based surveys by contact mode, where the check marks indicate which sampling strategies are mainly associated with which contact methods.

Contact Method		***Sampling Strategy***							
		Probability-Based					**Non-Probability-Based**		
		List-based sampling frames	Non-list-based random sampling	Systematic sampling	Mixed mode survey with Internet-based option	Pre-recruited survey panel	Entertainment polls	Unrestricted self-selected surveys	Volunteer (opt-in) panels
Internet-based	Web			✓			✓	✓	✓
	E-mail	✓							
Non-Internet-Based	Telephone	✓	✓		✓	✓			
	Postal Mail	✓			✓				
	Other: TV, print advertising, etc.						✓	✓	✓

Table 3. *Sampling strategies* for Internet-based surveys *by contact mode.*

Note that we are focusing explicitly on Internet-based surveys in Table 3. So, for example, while systematic sampling can be applied to phone or mail

DRAFT

surveys, the telephone is not likely to be used as a contact medium for an Internet-based survey using systematic sampling and hence those cells in the table are not checked. Similarly, while there is a plethora of phone-in entertainment polls, neither the telephone or postal mail are used to contact respondents to take Internet-based entertainment polls.

From Table 3 we can broadly summarize the application of the various Internet-based survey methods and their limitations as follows.

- *Entirely web-based surveys, meaning surveys in which the potential respondents are contacted on the web and take a web survey, are chiefly limited to collecting data from non-probability-based samples.*
    - The exception is systematic sampling for pop-up/intercept surveys that are predominantly used for customer satisfaction types of surveys associated with specific websites or web pages.
    - Respondent contact for Internet-based surveys using non-probability samples can also be conducted via traditional (non-Internet-based) media and advertising.
    - Non-probability samples have various research uses but are inappropriate for statistically generalizing beyond the particular convenience sample.
- *Research which requires generalizing the survey results to some larger population (in other words, surveys that require probability sampling) is very limited when using an Internet-based contact mode (web and e-mail).*
    - E-mail is only useful as a contact mode if a list of e-mail addresses is available. Such a list is an actual or *de facto* sampling frame, from which a sample may be drawn or a census attempted.
    - The population of inference is usually quite limited when using an e-mail address list sampling frame, assuming a probability sample is drawn. It is generally the sampling frame itself.
    - A poorly conducted census of the entire e-mail list may limit the survey results even further since nonresponse and other biases may preclude generalizing even to the sample frame.

DRAFT

- *If the research objectives require generalizing the survey results to some larger, general population, then respondents will most likely have to be contacted by some non-Internet-based medium.*
    - If the population of inference is a population in which some of the members do not have e-mail/web access, then the contact mode will *have* to be a non-Internet-based medium.
    - Under such conditions, the survey will have to be conducted using a mixed-mode so that those without Internet access can participate. Conversely, lack of a non-Internet-based survey mode will result in coverage error with the likely consequence of systematic bias.
    - Pre-recruited panels can provide ready access to pools of Internet-based survey respondents, but to allow generalization to some larger, general population such panels need to be recruited using probability sampling methods from the general population (usually via RDD). And, even under such conditions, researchers need to carefully consider whether the panel is likely to be subject to other types of bias.

DRAFT

Further Reading

- *Mail and Internet Surveys: The Tailored Design Method* by Dillman and *Survey Errors and Survey Costs* by Groves. Each of these texts focuses on the entire process of designing and fielding surveys, not just sampling. *Mail and Internet Surveys* addresses both traditional survey modes and well as Internet-based survey modes. *Survey Errors and Survey Costs* provides a complete structure and taxonomy for describing survey errors and a useful framework for thinking about how to reduce the total survey error.
- *Conducting Research Surveys via E-Mail and the Web* by Schonlau, Fricker, and Elliott 'is a practical and accessible guide to applying the pervasiveness of the internet to the gathering of survey data in a much faster and significantly less expensive manner than traditional means of phone or mail communications. Yet online surveys have their own pitfalls that can adversely skew data results. Individual chapters cogently address how to take maximum advantage of the Internet while preserving data integrity, and provide an invaluable look at the future of internet-based survey methods.' (Midwest Book Review, 2003)
- 'Review: Web Surveys: A Review of Issues and Approaches' by Mick P. Couper, published in *The Public Opinion Quarterly*, is an excellent and highly-cited article that emphasizes many of the points and ideas discussed in this chapter. Though written in 2000, it is still current and relevant. It also provides additional examples to those presented in this chapter.
- *Sampling Techniques* by Cochran is one of the classic texts on the mathematical details of survey sampling, covering a wide range of sampling methods applicable to all types of survey efforts. It is an excellent reference for those who must design a survey sampling plan. However, as a reviewer on Amazon.com wrote, 'it is not for the faint hearted or those who do not have a bent towards statistics or numerical analyses.'

DRAFT

References

Alvarez, R.M., Sherman, R.P., and VanBeselaere, C. (2002) 'Subject Acquisition for Web-Based Surveys', dated September 12, 2002, accessed on-line at survey.caltech.edu/alvarez.pdf on September 29, 2006.

Andrews, D., Nonnecke, B., and Preece, J. (2002) 'Electronic Survey Methodology: A Case Study in Reaching Hard-to-Involve Internet Users', *International Journal of Human-Computer Interaction*, 12(2): 185-210..

Berson, I.R., Berson, M.J., and Ferron, J.M. (2002) 'Emerging Risks of Violence in the Digital Age: Lessons for Educators from an Online Study of Adolescent Girls in the United States', *Journal of School Violence*, 1(2), 51-72. Published simultaneously online in *Meridian, A Middle School Computer Technologies Journal* accessed at www2.ncsu.edu/unity/lockers/project/meridian/sum2002/cyberviolence/cyberviolence.pdf on September 29, 2006.

Cochran, William G. (1977) *Sampling Techniques.* New York: John Wiley & Sons.

Comley, P. (2000) 'Pop-up Surveys: What Works, What Doesn't Work and What Will Work in the Future.' Proceedings of the ESOMAR worldwide Internet conference Net Effects 3, volume 237. Amsterdam, NL: ESOMAR. Accessed at www.virtualsurveys.com/news/papers/paper_4.asp on September 21, 2006.

Couper, Mick P. (2000) 'Review: Web Surveys: A Review of Issues and Approaches', *The Public Opinion Quarterly,* 64(4): 464-494.

Couper, Mick P., J. Blair, T. Triplett. (1999) 'A Comparison of Mail and E-Mail for a Survey of Employees in Federal Statistical Agencies.' *Journal of Official Statistics*, 15(1): 39-56.

Dillman, D.A. (2000) *Mail and Internet Surveys: The Tailored Design Method*, second edition. New York: John Wiley & Sons.

Dillman, D.A., Tortora, R.D. and Bowker, D. (1999) 'Principles for Constructing Web Surveys', accessed online at www.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf on January 27, 2005.

Fahmy, S., Fosdick, S.B., and Johnson, T.J. (2005) 'Is Seeing Believing? A Survey of Magazine Professionals' Practices and Attitudes Toward Ethical Standards for Photographs', *Journal of Magazine and New Media Research,* Spring issue, 1-16. Accessed at www.bsu.edu/web/aejmcmagazine/journal/archive/Spring_2005/Article%201.pdf on September 21, 2006.

Fink, Arlene (2003) *How to Sample in Surveys, 2nd ed.* The Survey Kit, volume 7. Thousand Oaks: Sage Publications.

Fowler, Jr., Floyd J. (2002) *Survey Research Methods, 3rd ed.* Applied Social Research Methods Series, volume 1. Thousand Oaks: Sage Publications.

Fricker, Jr., R.D., and Schonlau, M. (2002) 'Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature', *Field Methods*, 14: 347-367.

Fricker, S., Galesic, M., Tourangeau, R., and Yan, T. (2005) 'An Experimental Comparison of Web and Telephone Surveys', *Public Opinion Quarterly*. 69(3): 370-392.

Göritz, A.S. (2006) 'Incentives in Web Studies: Methodological Issues and a Review', *International Journal of Internet Science*. 1(1): 58-70.

Groves, Robert M. (1989) *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

Jackob, N., Arens, J., and Zerback, T. (2005) 'Sampling Procedure, Questionnaire Design, Online Implementation, and Survey Response in a Multi-national Online Journalist Survey', paper presented at the Joint WAPOR/ISSC conference Conducting International Social Surveys.

DRAFT

Accessed online at ciss.ris.org/uploadi/editor/1132070316WAPORPaper.pdf on September 30, 2006.

Kish, Leslie (1965) *Survey Sampling.* New York: Wiley-Interscience. New edition (February 6, 1995).

Krishnamurthy, S. (2002) 'The Ethics of Conducting E-mail Surveys', *Readings in Virtual Research Ethics: Issues and Controversies,* accessed online at ssrn.com/abstract_id=651841 on September 30, 2006.

Kypri, K., and Gallagher, S.J. (2003) 'Incentives to Increase Participation in an Internet Survey of Alcohol Use: A Controlled Experiment', *Alcohol & Alcoholism.* 38(5): 437-441.

Lockett, A., and Blackman, I. (2004) 'Conducting Market Research Using the Internet: The Case of Xenon Laboratories', *Journal of Business & Industrial Marketing*, 19(3), 178-187.

Midwest Book Review. (2003) 'Book Review: Take Advantage of the Internet and Preserve Data Integrity' accessed online at http://www.amazon.com/Conducting-Research-Surveys-E-Mail-Web/dp/0833031104/sr=1-3/qid=1160191573/ref=sr_1_3/102-5971652-1424906?ie=UTF8&s=books on October 6, 2006.

Pineau, V. and Dennis, J.M. (2004) 'Methodology for Probability-Based Recruitment for a Web-Enabled Panel', dated November 21, 2004. Downloaded from www.knowledgenetworks.com/ganp/reviewer-info.html on July 21, 2006.

Ross, I. (1977) *The Loneliest Campaign: The Truman Victory*, reprint edition, Greenwood Press. Accessed online at www.trumanlibrary.org/whistlestop/study_collections/1948campaign/large/docs/documents/index.php?documentdate=1968-00-00&documentid=53&studycollectionid=Election&pagenumber=1 on October 1, 2006.

Schonlau, Matthias, Zapert, K., Simon, L.P., Sanstad, K.H., Marcus, S.M., Adams, J., Spranca, M., Kan, H., Turner, R., and Berry, S.H.. (2004*)* 'A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey', *Social Science Computer Review*, 22(1): 128-138.

Schonlau, Matthias, Fricker, Jr. Ronald D., and Elliott, Marc N. (2002*) Conducting Research Surveys via E-Mail and the Web*, MR-1480-RC, Santa Monica: RAND.

Sheehan, K.B. (1999) 'Using E-mail to Survey Internet Users in the United States: Methodology and Assessment', *Journal of Computer Mediated Communication*, (4)3. Accessed online at http://jcmc.indiana.edu/vol14/issue3/sheehan.html on July 6, 2006.

Shillewaert, N., Langerak, F., and Duhamel, T. (1998) 'Non-probability Sampling for WWW Surveys: A Comparison of Methods', *Journal of the Market Research Society*, 40(4): 307-322.

Siah, C.Y. (2005) 'All that Glitters is Not Gold: Examining the Perils and Obstacles in Collecting Data on the Internet', *International Negotiation*, 10: 115-130.

Soetikno, R.M., Provenzale, D., and Lenert, L.A. (1997) 'Studying Ulcerative Colitis over the World Wide Web', *American Journal of Gastroenterology*, 92(3): 457-460.

Squires, P. (1988) 'Why the 1936 *Literary Digest* Poll Failed', *The Public Opinion Quarterly,* 53(1): 125-133.

Witte, J.C., Amoroso, L.M., and Howard, P.E.N. (2000) 'Research Methodology: Method and Representation in Internet-Based Survey Tools – Mobility, Community, and Cultural Identity in Survey2000', *Social Sciences Computer Review*, 18(2): 179-195.

Wright, T., and Tsao, H.J. (1983) 'A Frame on Frames: An Annotated Bibliography', in T. Wright (ed.), *Statistical Methods and the Improvement of Data Quality.* New York: Academic Press.

DRAFT

Zetterberg, H.L. (2004) 'US Election 1948: The First Great Controversy about Polls, Media, and Social Science', paper presented at the WAPOR regional conference on 'Elections, News Media and Public Opinion' in Pamplona, Spain, November, 24-26. Accessed online at http://www.zetterberg.org/Lectures/l041115.htm on October 1, 2006.

DRAFT